OMB No. 3137-0035
expires 12/31/2000

1999 IMLS National Leadership Grants — *Application Form* **6.3**

LL-90178

Face S. University of California, Berkeley

1. Applicant: Regents of the University of California

2. Applicant's Mailing Address: c/o Sponsored Projects Office
336 Sproul Hall

3. City: Berkeley

4. State: CA

5. Zip Code: 94720-5940

6. Name and Title of Authorizing Official: Lynn E. Deetz

7. Business Phone of Authorizing Official: (510) 643-6113

8. Name of Project Director: Michael K. Buckland

9. Business Phone of Project Director: (510) 642-3159

10. FAX Number of Applicant: (510) 642-8236

11. e-mail address of Project Director: buckland@sims.berkeley.edu

12. Sponsoring institution/parent organization, if applicable (e.g., municipality, state, or university)
[X] check if this entity will manage funds if an award is made. Name and address: Berkeley, CA 94720-5940
Regents of the University of California, c/o Sponsored Projects Office, 336 Sproul Hall

13. Governing control of applicant [2]* (turn page for selections) * if 6, please specify _____

14. Type of organization [13]* (turn page for selection) select only one * if 14, please specify _____

15. Employer identification number 946002123

16. Type of project (turn page for selections) [2] select only one

17. Use of technology: [x] check box if application proposes the use of hardware and/or software as a significant aspect of the project.

18. GRANT AMOUNT REQUESTED $ 232,825 .00

19. Amount of Matching Funds $ 122,590 .00

20. Grant Period (Starting Date) 10 / 01 / 99 — 09 / 30 / 2001 (Ending Date)

21. Identify other Federal agencies that either have contributed support or have a pending proposal for support of these project activities and indicate the amount of support contributed or requested.

Name of Agency	Contributed/Pending	Amount
NONE		

22. In the space below, include the names of any organizations that are official partners of the project.

23. Certification: Lynn E. Deetz CFA
Signature of Authorizing Official — Date: 3-18-99

2. Abstract.

SEAMLESS SEARCHING OF NUMERIC AND TEXTUAL RESOURCES

The dream for the use of new technology in libraries is to support seamless searching across an increasing range of resources on a growing digital landscape. The reality is that network-accessible digital resources, like the contents of a well-stocked reference library, are quite heterogeneous, especially in the variety of indexing, classification, categorization, and other forms of "metadata."

The contribution of this project would be to demonstrate *improved access to written material and numerical data on the same topic* when searching two quite different kinds of database: text databases (books, articles, and their bibliographic records) and numerical data (socio-economic databases). The problem with doing this type of search is that there has, until now, been no easy path to integrate numeric databases with bibliographic and textual databases which might contain knowledge about cause and effect. The vocabulary which classifies the numeric data may be quite different from the subject headings used for books, magazine articles, and newspaper stories about the same topic of interest. Also there needs to be an environment of search support that facilitates such transverse searching, establishing connections, transferring data and invoicing appropriate utilities in a helpful way.

This project addresses both problems in two phases:

Phase I is the development and demonstration of a library gateway providing search support for searching *both* text *and* socio-economic numeric databases. If you have a query, the gateway would help you do a search in each type of database. The Phase I gateway would accept a query in the library users' own terms and would suggest what terms in the specialized categorization used in the resource to be searched.

Phase II is demonstration of a library gateway supporting searches *between* text and numeric databases. If you found some thing interesting in a socioeconomic database, the gateway would help you to find documents on the same topic in a text database - and vice versa.

Assistance in selecting the best search terms in the target database is made possible by the use of "Entry Vocabulary Modules," which resemble Melville Dewey's "Relativ Index," but are created using statistical association techniques developed at Berkeley.

3. Narrative.

SEAMLESS SEARCHING OF NUMERIC AND TEXTUAL RESOURCES

Introduction and National Impact

There has been a massive investment worldwide in the hardware, software, and protocols to enable remote access to a wide variety of resources. Libraries are now in a network environment. The central purpose of libraries is to provide access to information. New technology allows and requires innovation to achieve that central purpose (Buckland, 1992). The possibility of enabling library users to access to a widening range of digital resources is an exciting opportunity. One consequence has been the reorientation of library technical services and library systems development from creating catalogs towards creating "gateways" to more varied sources (Norgard et al., 1993).

The dream is to support seamless searching across an increasing range of resources on a growing digital landscape. The reality is that network-accessible digital resources, like the contents of a well-stocked reference library, are quite heterogeneous, especially in the variety of indexing, classification, categorization, and other forms of "metadata." There is increasing difficulty for library users when searching, because the number of unfamiliar, remote repositories is increasing and, with them, the amount and proportion of unfamiliar metadata vocabularies is increasing. Underutilization of technically accessible resources, decreased search effectiveness, and a poor return on investments result.

Building a more seamless service comes one step at a time. The contribution of this project would be to demonstrate *improved access to written material and numerical data on the same topic* when searching two quite different kinds of database: text databases (books, articles, and their bibliographic records) and numerical data (socio-economic databases).

Socio-economic datasets as an example. In past decades the Federal census and other similar socio-economic data could be found in printed form in large and specialized libraries. The printed volumes fitted the norms of library provision. One could wander from one part of the library to another, assembling pertinent evidence for the term paper, article, memorandum, or other purpose. But that convenient, if limited, scenario has been steadily changed by advances in technology. Increasingly socio-economic data are available in--often only in--electronic form and moved away from traditional norms of library service. "Data archives" designed to house large stores of socio-economic numeric data and to provide technical support for access and use of such material grew up separately from libraries. The cataloging of socio-economic datasets has, in general, not followed neither library practices nor the standards adopted for online bibliographic systems. The provision of access to numeric data and the provision of access to textual resources have gone their separate ways.

From the library user's perspective, it is clear that this unhelpful technical, organizational, and bibliographic separation needs to be bridged. Whenever anyone reads an article on any topic, it is reasonable to wonder whether relevant quantitative data exist. Whoever finds intriguing statistical data may wonder whether newspapers, journals or books have discussed this phenomenon.

The project proposed here is in two phases:
Phase I is the development and demonstration of a library gateway providing search support for searching *both* text *and* socioeconomic numeric databases. If you have a query, the gateway would help you do a search in each type of database. The Phase I gateway would accept a query in the library users' own terms and would suggest what terms in the specialized categorization used in the resource to be searched. This latter capability is sometime called an "Entry Vocabulary Module" since it starts with the vocabulary the searcher brings to the search.
Phase II is demonstration of a library gateway supporting searches between text and numeric databases. If you found some thing interesting in a socioeconomic database, the gateway would help you to find documents on the same topic in a text database - and vice versa.

Example: From text to numbers

Suppose you wish to investigate the effects of mad cow disease on the imports of beef to the United States from Britain and found articles about the U.S. ban on imports of beef from the newspaper articles. e.g.

"U.S. bans import of most European meat". *Los Angeles Times*, vII6, nl4 (Dec 14,1997):A22. (On fear of mad cow disease.)

"Ban on cattle and sheep is extended to all Europe." *New York Times*, vl47, sec 1, (Dec 14,1997): 16(N), 42(L). (The U.S. Agriculture Department responds to threat of 'MadCow' disease).

Such writings are unlikely to include exact historical numbers as to the amount of beef cattle imports from Britain to the U.S. So what are the facts? Data are available at an Internet site http://govinfo.kerr.orst.edu/import/import.html which has U.S. annual imports totals extracted from CD-Rom. A search on that source will yield some surprising numbers: There has been *no* report edible beef imports from the United Kingdom, but that there has been significant numbers of edible meat from animals other than beef, e.g. frozen hams and deer meat.

Example: From numbers to text

An example in the other direction could be noticing a sudden increase of imports into the USA through Los Angeles of shrimp and prawn from Vietnam and curiosity about the political background and economic consequences. The numbers show:

--- U.S. Imports of Merchandise ---
General Imports: Imports for Consumption

Year	Quantity	Customs Value
SHRIMP/PRAWN SHELL-ON COUNT SIZE 33-45 PER KG FRZN		
(HS: 0306130006) (SIC: 0913)		
Unit of Quantity -- Kilograms		
FROM: Vietnam	THRU: LOS ANG	
1993	0	0
1994	48,782	676,930
1995	247,707	3,520,806
1996	562,427	7,864,052

Taking the keywords "Import" and "Vietnam" over to an online bibliographic databases of newspaper articles retrieves, among others,

Iritani, Evelyn. “Normalizing ties to Vietnam important steps for U.S. firms; California stands to profit handsomely when barriers fall to trade with fast-growing country.” *Los Angeles Times*, v114, (July 12, 1995): D1.

"Hanoi's trade deficit." *New York Times*, v143, (July 15, 1994): D15(L). (Vietnam imports increasing faster than exports).

The problem with doing this type of search is that there has, until now, been no easy path to integrate numeric databases with bibliographic and textual databases which might contain knowledge about cause and effect. Two significant problems are encountered when seeking to traverse a search across from libraries textual resources to numeric databases or vice versa:

1. The vocabulary which classifies the numeric data may be quite different from the subject headings used for books, magazine articles, and newspaper stories about the same topic of interest. For example, searching for Federal imports data for "automobiles" returns no results, even though billions of dollars of U.S. foreign exchange goes into auto imports. Searching under "car" yields data on railroad and tramway rolling stock. To get automobiles, the searcher needs to know to search under "P" for "Passenger Motor Vehicle."

2. There needs to be an environment of search support that facilitates such transverse searching, establishing connections, transferring data and invoking appropriate utilities in a helpful way.

This proposal addresses both problems. It is research and demonstration designed to improve access to library and information resources. It extends the most recent research in library science to demonstrate and test a potential solution to a neglected real-world problem: How to support searches for both numeric and textual databases for information on the same topic. By demonstration, evaluation, description, and a www-accessible prototype, any librarian with access to the Internet will be able to try out this solution. By the free distribution of software all library system developers will be encouraged to adopt and adapt what is developed. The research and the prototype to be made available to librarians and library system developers to test-drive

will necessarily focus on a few carefully chosen resources, but it will be clear that the same techniques are generally applicable to other online resources.

This proposal advances two of the three IMLS priorities for Research and Demonstration projects: It involves research and demonstration to enhance library services through effective and efficient use of new and appropriate technologies; and it will enhance the ability of library users to make more effective use of information resources. This narrative explains the question, the plan of work, the solution to be demonstrated and the evaluation procedures.

Adaptability.

The problem addressed is a universal, not a local one. Access will be demonstrated to well-known and widely used databases. Once an effective prototype capable of supporting user-friendly traversing from bibliographical and numeric databases has been developed, made available, and becomes known, we fully expect that this kind of functionality will come to be a generally expectation of library services. What works in the proposed prototype gateway should, in principle, work anywhere.

We are committed to designing a prototype that other developers will want to adopt and adapt for their own purposes. In order to allow widespread adoption, the prototype is to be designed for general adoption using prevailing standards: SGML, WWW-access, Z39.50 client and server, and clear, explicit documentation. (See Design below). We avoid (as far as we can) proprietary or specialized software for this reason. As with our other work, the demonstration prototypes will be made available for open WWW access. As we have done before, software will be made available without charge on an ftp server. Some developers prefer to create their own production software:
Access to the project software will help them by showing in detail how it can be done.

Design.

The design of the project is in two Phases. Each Phase has two parts: (i) Research and demonstration and (ii) Evaluation.

Phase 1: *A library gateway for searching **in** both text and socioeconomic databases:* An open WWW-accessible site will provide a gateway with the following amenities:

1. An invitation to searchers to enter their query in their own terms in a dialog box;
2. A menu from which to select the database (text or numeric) they wish to search;
3. An Entry Vocabulary Module (explained more fully below) which takes the query, maps it against the subject headings (or other metadata vocabulary) in the database selected, and provides the searcher with a prompting of the terms most likely to match their query in the target database;
4. Where the target database has a hierarchical structure (e.g. the INSPEC Thesaurus and the Library of Congress Classification), the Entry Vocabulary will offer support for navigating within that structure in case some related term fits better; and
5. When the searcher is ready, the gateway passes the query on to the selected database.

The socioeconomic numeric databases intended are the *U.S. Exports of Merchandise on CD-ROM (Monthly Series), U.S. Imports of Merchandise on CD-ROM (Monthly Series), County Business, Patterns, 1995: U.S. Summary, State and County Data, and Census of Population and, Housing, 1990: Equal Employment Opportunity File*. The intended text / bibliographic databases include the MELVYL online library catalog, the *L.A. Times*, *INSPEC, BIOSIS*, and one or more

indexing and abstracting services to be determined.

The Phase I gateway is then evaluated as described on the Evaluation section below.

Phase II: *A library gateway supporting searches **between** text and numeric databases.* If you found some thing interesting in a socioeconomic numeric database, the gateway would help you to find documents on the same topic in a text database-and vice versa. The Phase II gateway is the same as the Phase I gateway except that data found in a database can form the basis for a search in another database. In the examples above the bibliographic records for the articles (including title words) could form the basis for a search in the numeric database. For the numeric data, the column and row labels (e.g. "Imports," "Shrimp," and "Vietnam") can be used to search in bibliographic databases. The specialized metadata such as the International Harmonized Commodity and Coding System (HS0306130006) and the Standardized Industrial Classification (SIC 0913) can be used or expanded into English terms to form a query, which a searcher can use or amend for keyword searches or, through use of an Entry Vocabulary Module, translate into the specialized terms or subject codes of the target database.

The Phase II gateway will also be evaluated as described below.

A Note on Entry Vocabulary Modules: All indexes, thesauri, classification, and categorization schemes are more or less stylized and specialized. Experienced searchers know that some familiarity with the indexing or categorization is needed for effort-effective searching. Help can be provided: Melville Dewey provided a "Relativ Index" to his Decimal Classification. Using his reformed spelling, Dewey wrote: "This alfabetic Index, the most important feature of the sistem, consists of headings gatherd from a great variety of sources, as uzers of the sistem hav found them desirabl... The Index givs similar or sinonimus words.... so any intelijent person wil surely get the ryt number ..... (Olding, 1966, 82-91).

A "Relativ Index" can be helpful when searching any unfamiliar terminology, but is very expensive to create. A research and demonstration project entitled "Search Support for Unfamiliar Metadata Vocabularies," funded by the Defense Advanced Research Projects Agency (DARPA) has enabled the three Principal Investigators to develop and demonstrate how a form of "Relativ Index" known as Entry Vocabulary Modules can be created using statistical and computational techniques. These computer-generated Relativ Indexes are currently available for BIOSIS Concept Codes, Library of Congress Classification science scheduled, the U.S. Patent & Trademark Office's Patent Classification, the WIPO International Patent Classification, the INSPEC Thesaurus, and the Standard Industrial Classification. Reviewers are encouraged to try using these openly WWW-accessible prototype Entry Vocabulary Modules and, in doing so, see how prototypes can be made openly web-accessible). The project website is at http://www.sims.berkeley.edu/research/metadata/. The prototypes are at http://www.sims.berkeley.edu/research/metadata/oasis.html. A short, readable introduction in *Dlib Magazine* is at http://www.dlib.org/dlib/january99/buckland/Olbuckland.html. We seek to demonstrate how that Federally-funded research can be useful in libraries.

Management Plan.

Responsibility for the negotiation of grants is in the hands of the campus Sponsored Projects Office. Budgetary and administrative procedures will be carried out in the School of Information

Management & Systems. Michael Buckland, as Principal Investigator, will be responsible for the conduct of the project. All of the above must follow detailed University and Federal regulations.

Prof. Buckland has had extensive experience in academic administration, library administration, and in the management of research and development projects.

The project has design and timetable is quite simple and straightforward:
Task 1: Technical specifications for the Phase I Gateway: Finalize software, hardware, standards and selection of databases, 10/99 - 11/99.
Task 2: Programming, technical development, and testing of Gateway 1. 12/99 - 6/00.
Task 3: Evaluations of Gateway 1, 7/00 - 9/00.
Task 4: Technical specifications for the Phase 11 Gateway: 10/00.
Task 5: Programming, technical development, and testing of Gateway H. 11/00 - 5/01.
Task 6: Evaluations of Gateway H, 6/01 - 8/01.
Task 7: Completion: Complete documentation and ensure continuing software availability, 9/01.

The project team will meet weekly, issue quarterly progress reports, maintain an informative project website. Dissemination through writings and presentations is an ongoing activity throughout.

Budget.

The staffing budget is the best estimate of the three P.I.s, based on their experience with several years of similar prior research. Details are in the budget and budget narrative. The development of prototypes suitable for public access on the WWW requires more and more robust software development and testing than prototypes for only in-house use. If additional time commitment by the P.I. and Co-P.I.s proves to be needed, it will be contributed as additional matching.

The budgeted matching slightly exceeds 33% of the total. It understates the actual match since all equipment needs (considerably more that the one workstation listed) will be contributed.

Personnel.

The resumes of the Michael Buckland, Fredric Gey, and Ray Larson are attached. Each has substantial appropriate experience and expertise in the specific kind of work proposed. They have usefully complementary backgrounds and are used to working together as a team.

Professor Buckland has been actively involved in library research and demonstration projects since 1967 and has performed almost all aspects of library service from stack page to multi-campus library coordination. He was responsible for developing the award-winning Library Research Unit at the University of Lancaster (U.K.), for supervising the Library Systems Unit and the Library Management Research Unit when Assistant Director for Technical Services at Purdue University, and for development of the MELVYL system when Assistant Vice President for library Plans and Policies in the University of California system-wide administration. His nearly 900 publications include ***Redesigning Library Services*** (ALA, 1992), a best seller re-published in three foreign languages. Professor Buckland is a Life Member of the American Library Association and Immediate Past-President of the American Society for Information Science. The present proposal is a direct continuation of his long-term OASIS research program ("Otlet's Adaptive Search Information Service"), started in 1990 with support from the HEA-11 program, and concerned with making online searching in libraries easier and more powerful. Summary at:
http://www.sims.berkeley.edu/research/oasis/

Dr. Fredric Gey, Assistant Director of UC Data Archive & Technical Assistance at the University of California, Berkeley (http:/Iucdata.berkeley.edu). He has substantial experience in designing improved access to numeric databases, including, in collaboration with the campus libraries, a digital library of 135 gigabytes of U.S. federal statistics (primarily 1990 Census Data), which is now in use at the U.S. Census Bureau's web site, as well as detailed subject access from1990 census files for Hispanic and Asian populations which is available to the www from the UCBerkeley Sunsite http://sunsite.berkeley.edu/GovData/info (See Merrill, Parker, Gey, and Stuber, 1995). Dr. Gey's research includes text retrieval: He is principal investigator for National Science Foundation Grant IRI-9630765 (1996-1999) *Probabilistic Retrieval of Full Text Document Collections using Logistic Regression.*

Prof. Ray Larson worked on the development of the University of California NIELVYL system and has taught information retrieval and library automation at Berkeley since 1985. The CHESHIRE system, being developed by Prof. Ray Larson, is a "next-generation" online catalog and full-text information retrieval system using multiple advanced IR techniques to overcome twin problems of topical searching in online catalogs, search failure and information overload. It incorporates a client/server architecture with implementations of current information retrieval standards including Z39.50 and SGML. It is currently being deployed in the science libraries of the Berkeley campus, at the Federal Bureau of Investigation, and at Colchester and Liverpool in England its use and acceptance by local library patrons and remote network users are being evaluated. CHESHIRE is evidence of Professor Larson's expertise in crafting advanced library access. See http://cheshire.lib.berkeley.edu/

Doctoral students who specializing in precisely this kind of work and who have co-authored relevant technical papers would be hired for this project.

Evaluation.

Three formal, quantitative evaluations are proposed:

1. The most sophisticated technique available for evaluating search systems is that developed by the National Institute of Standards and Technology (NIST) for the successive Text Retrieval Conferences (TREC). Queries created by NIST will be used to search a typical text database (articles from the Los Angeles Times) with and without use of an Entry Vocabulary Module. The judgements already made by NIST staff on the relevance of individual documents to each query provide a impartial basis for evaluating the effectiveness of the Gateway and Entry Vocabulary Module in improving retrieval performance.

2.For an objective evaluation when NIST queries and relevance judgements are not available, titles of articles and of books that are descriptive of their topic will be used as queries to test the ability of the Entry Vocabulary Modules to predict which subject heading or classification was assigned. This technique was pioneered by Professor Larson. (See Larson 1992).

3. In addition, a more subjective, user-oriented evaluation will be used. Under controlled conditions, users' queries will be submitted two ways: directly and through the Gateway with its Entry Vocabulary Modules. The users will be asked to compare the two results (without knowing which is which) and to determine which one they prefer. Their preferences will determine whether

and how far the Gateway improves retrieval performance. (Sometimes called the Saracevic-Kantortechnique. See Saracevic & Kantor, 1988)

Dissemination.

Dissemination will be through multiple channels:

- an informative project website;
- demonstration gateways freely web-accessible;
- technical reports;
- papers presented at conferences and published in conference proceedings; and
- articles in leading professional and technical journals.

The dissemination will be crafted to reach audiences of librarians, of numeric data specialists, and of systems developers. Each of the three P.I.s already has a well-established record of perforating all of these dissemination activities.

Contributions and Matching Funds.

Faculty members have limited resources to contribute other than their own time. Professors Buckland and Larson are contributing 10% of their time during the academic year and Dr. Gey 5% year-round. This contribution will be increased should the project need it. The School is contributing a powerful workstation as a designated contribution (as well as all other hardware). Ex Libris Ltd is the vendor of the innovative ALEPH library systems software used by more than 450 library systems in 38 countries. Ex Libris Ltd will assign a twenty-five working days a year of the time of senior research and development staff as a contribution in kind. We have assessed this contribution conservatively at $10,000 in annual salary and fringe benefits-plus indirect costs.

These contributions total about 35% of total project cost.

The project will benefit from ongoing related research funded from other sources and from institutional support from our existing infrastructure of hardware, software, and access to databases, but these contributions are difficult to estimate in dollar terms and so are not in the budget.

Sustainability.

Dr. Gey is participating in this project because he intends to implement the result at UCData, the leading repository of numeric databases in California. Prof. Larson commits to incorporating the research into the CHESHIRE system which has been adopted in libraries in the USA and the U.K. Ex Libris Ltd is participating because it expects the project to add to the functionality of library automation systems. Prof. Buckland's prior work has been incorporated in to the University of California's MELVYL system.

Technical Knowledge.

Knowledge of best practices: The resumes attached indicate that the Michael Buckland, Fredric Gey, and Ray Larson each have some twenty years of experience in closely related research and development. All three have also had responsibility for the introduction and operation of innovative information services. All three have been awarded highly competitive research funding. Gey and Larson has been participants in NIST TREC conference, the international forum for testing innovative advanced text retrieval techniques.

Most innovative or appropriate technologies: The project proposes to deploy the latest relevant

research, funded by the Defense Advanced Research Projects Agency, and the latest open standards.... testing: The evaluation involves three different formal evaluations: That of the TREC conference; and objective test using tides as queries; and the "Saracevic-Kantor" approach to user evaluation.
Commitment to sharing: The prior research, development by the three P.I.s have adopted or adapted as follows: Buckland's OASIS techniques in support of adaptive searching have been incorporated in the enhancement of the University of California MELVYL system; Larson's CHESHIRE system has been made available without charge at Berkeley, at the FBI, and, in the U.K, at Liverpool and the Colchester; Gey's innovations have been implemented at Berkeley and the Census Bureau. All three have an extensive record of sharing their knowledge through publication, conference presentations, website content.

Information Access

This is a research and demonstration project in an academic department of a Land Grant University. The applicants are fully committed to communicating the increased availability of electronic access to specific audiences and to the general public to the extent that they are in a position to do so. Dr. Gey, as Assistant Director of University of California Data Archive & Technical Assistance (UCData), is responsible for an operational information service and he is fully committed to implementing the results of this project on a permanent basis. Prof. Larson is committed to incorporating the results into the CHESHIRE system, and advanced experimental system in operational use. The applicants see adoption of their work by others in operational systems as the justification for their efforts.

One More Example - and Conclusion.

If access to socio-economic datasets and text databases were more convenient to use, they would be used more. For example, the decline in the logging industry and loss of employment in northern California is a widely known and widely discussed. But if one looked in the federal County Business Patterns (available at http://fisher.lib.virginia.edu/cpb/home.html) one finds a startling seven-fold increase in wages in the forestry sector in Humboldt County, California, from 1991 to 1994:

County - Humboldt County -SIC 08 Forestry	Year	Annual Wages (in thousand $)
	1991	1,272
	1992	2,019
	1993	6,742
	1994	7,154
	1995	7,012

What's happening? A newspaper database using keywords "Humboldt" and "Logging" retrieves:

Curtius, Mary. The parting of the ‘Redwood curtain’; after the decline of logging and fishing, the North Coast is reinventing itself as a place for tourists, entrepreneurs – and prisoners. (Del Norte, Humboldt,...). *Los Angeles Times,* 116(Dec 28, 1996):A1, c.1.

Anderson, Harry. The monied set storms Redwood country. (economics of logging and real estate speculation in Humboldt County, California) *Los Angeles Times*, 1O9(Aug.28,1990):D1, col.2.

It is the integration of these pieces from disparate resources into a unified whole, which we seek to achieve with project.

The dream is to support seamless searching across an increasing range on resources. This proposal advances two of the three IMLS priorities for Research and Demonstration projects: It involves research and demonstration to enhance library services through effective and efficient use of new and appropriate technologies; and it will enhance the ability of library users to make more effective use of information resources. This proposal contributes towards that dream by using the very latest research on metadata mapping to make searching easier and, specifically, to demonstrate support for seamless searching across both numeric and textual resources.

References

M. K. Buckland, B.A. Norgard, & C. Plaunt. Design of an adaptive library catalog. In *Networks, Telecommunications, and the Networked Information Revolution: Proceedings of the ASIS 1992 Mid-Year Meeting May 27-30,1992*, 165-71. Silver Spring, MD: American Society for Information Science, 1992.

M. K. Buckland. *Redesigning Library Services: A Manifesto*. Chicago: ALA, 1992. Http://sunsite.berkeley.edu/Literature/Library/Redesigning/html.html.

M. K. Buckland. From catalog to selecting aid. *ALCTS Newsletter 5*, no. 2 (1994): Insert A-D. (From catalog to gateway: Briefings from the CFFC, 2).

M. K. Buckland, M. H. Butler, B.A. Norgard & C. Plaunt. OASIS: A front-end for prototyping catalog enhancements. *Library Hi Tech Issue*, v.40 (1992):7-22.

M. K. Buckland, M. H. Butler, B.A. Norgard & C. Plaunt. Prototyping enhanced online search capability. In *National Online Meeting*, 14th, 1993. Proceedings. Ed. by M. Williams, 51-56. Medford, NJ: Learned Information, 1993.

R. R. Larson. 1991. Classification Clustering, Probabilistic Information Retrieval and the Online Catalog. *Library Quarterly*, vol. 61, no. 2, April 1991: 133-173.

R. R. Larson. Experiments in Automatic Library of Congress Classification. *Journal of the American Society for Information Science*, v. 43 no. 2, March 1992:130-148.

R. R. Larson, Jerome McDonough, Lucy Kuntz, Paul OLeary, and Ralph Moon. Cheshire 11: Designing a Next-Generation Online Catalog. *Journal of the American Society for Information Science*, v.47(7), July 1996:555-567.

D. Merrill, N. Parker, F. Gey, and C. Stuber. The University of California CD-ROM Information System. *Communications of the ACM* (Issue on Digital Libraries), April 1995:51.

B.A. Norgard, M.G. Berger, M. K. Buckland, & C. Plaunt. The online catalog: From technical services to access service. *Advances in Librarianship* v.17 (1993):111-148.

Office of Management and Budget. *Standard Industrial Classification Manual*. NTIS, 1987.

R. K. Olding, ed. 1966. *Readings in Library Cataloguing*. Hamden, CT: Archon Press.

T. Saracevic and P. Kantor. A study of information seeking and retrieving. *Journal of the American Society for Information Science.* v. 39 (3) May 1988: 161-216.

Schedule of Completion

SEAMLESS SEARCHING OF NUMERIC AND TEXTUAL RESOURCES

The project has design and timetable is quite simple and straightforward:

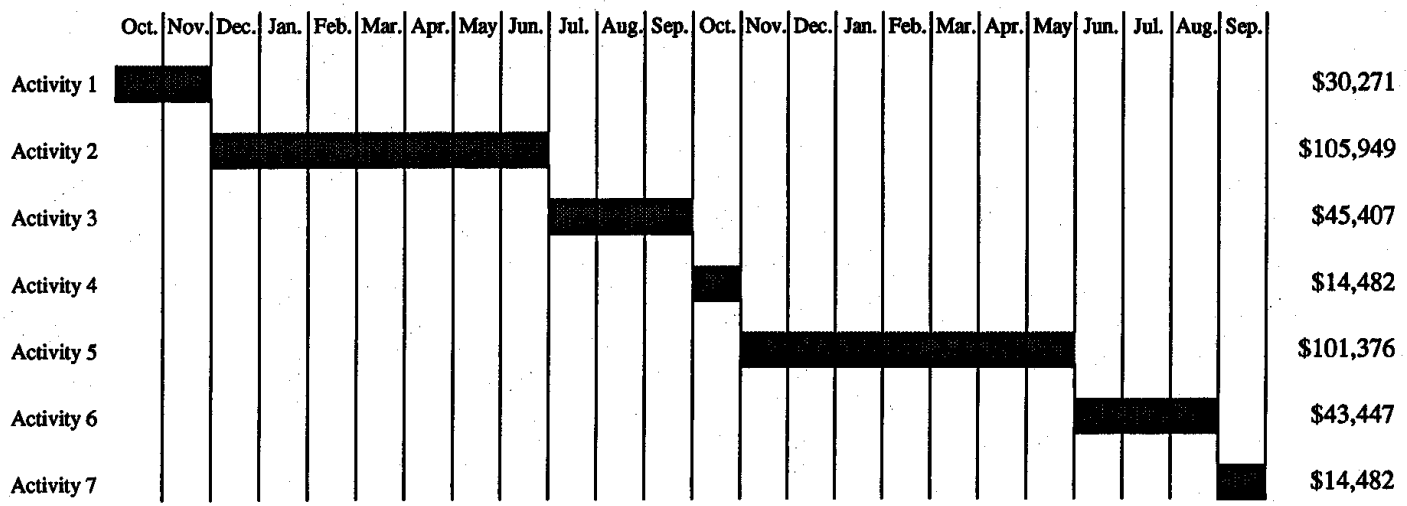
Activity 1: Technical specifications for the Phase I Gateway: Finalize software, hardware, standards and selection of databases, 10/99 – 11/99.
Activity 2: Programming, technical development, and testing of Gateway 1. 12/99 - 6/00.
Activity 3: Evaluations of Gateway 1, 7/00 - 9/00.
Activity 4: Technical specifications for the Phase H Gateway: 10/00.
Activity 5: Programming, technical development, and testing of Gateway H. 11/00 - 5101.
Activity 6: Evaluations of Gateway 11, 6/01 - 8/01.
Activity 7: Completion: Complete documentation and ensure continuing software availability, 9/01.

The project team will meet weekly, issue quarterly progress reports, maintain an informative project website. Dissemination through writings and presentations is an on going activity throughout.

See attached Gantt chart.

Schedule of Completion

IMLS Proposal: University of California, Berkeley. M. Buckland. Seamless Searching of Numeric and Textual Resources.
Activities. October 1, 1999 - September 30, 2001.



Activity 1	$30,271
Activity 2	$105,949
Activity 3	$45,407
Activity 4	$14,482
Activity 5	$101,376
Activity 6	$43,447
Activity 7	$14,482
	$355,415

Project Budget Form Front

SECTION 1: DETAILED BUDGET
YEAR 1

Name of Applicant Michael K. Buckland

SALARIES AND WAGES (PERMANENT STAFF)

NAME/TITLE	No.	Method of Computation	IMLS	Match	Total
		TOTAL SALARIES AND WAGES	**$15,791**	**$22,850**	**$38,641**

SALARIES AND WAGES (TEMPORARY STAFF HIRED FOR PROEJCT)

NAME/TITLE	No.	Method of Computation	IMLS	Match	Total
Graduate Student Researchers	(2)	see Budget Attachment A	35,160		35,160
		TOTAL SALARIES AND WAGES	**$35,160**	**$0**	**$35,160**

FRINGE BENEFITS

Rate	Salary Base	IMLS	Match	Total
	TOTAL FRINGE BENEFITS	**$30,160**	**$3,885**	**$34,045**

CONSULTANT FEES

Name/Type of Consultant	Rate of Compensation (Daily or Hourly)	No. of Days or Hrs) on Project	IMLS	Match	Total
		TOTAL CONSULTANT FEES			

TRAVEL

From/To	Number of Persons	Number of Days	Subsistence Costs	Transportation Costs	IMLS	Match	Total
West/East Coast Conf.	(4)	(50)	4,425	5,000	9,425		9,425
Research Meeting	(2)	(10)	250	974	1,224		1,224
			TOTAL TRAVEL COSTS		**$10,649**	**$0**	**$10,649**

Project Budget Form Back

SECTION 1 CONTINUED

YEAR 1

MATERIALS, SUPPLIES, AND EQUIPMENT

Item	Basis/Method of Cost Computation	IMLS	Match	Total
Copying, printing, computer supplies.		1,000	0	1,000
Dell Pentium II Workstation		0	10,771	10,771
TOTAL COST OF MATERIAL, SUPPLIES, & EQUIPMENT		**$1,000**	**$10,771**	**$11,771**

SERVICES

Item	Basis/Method of Cost Computation	IMLS	Match	Total
	TOTAL SERVICES			

OTHER

Item	Basis/Method of Cost Computation	IMLS	Match	Total
Effort contributed by Ex Libris		0	10,000	10,000
	TOTAL COST OF OTHER	**$0**	**$10,000**	**$10,000**

TOTAL DIRECT PROJECT COSTS	**$92,760**	**$47,506**	**$140,266**

INDIRECT COSTS

Select either item A or B and complete C.

Applicant is using

A. an indirect cost rate which doesnot esceed 20% of direct costs

or

B. an indirect cost rate negotiated with a Federal agency (copy attached)

Department of Health & Human Services — Name of Federal Agency

3/25/97 — Effective Date of Agreement

C. Rate base(s) Amount(s)

50.4% of $ 65,174

Amount: $32,848

TOTAL INDIRECT COSTS	**$32,848**

Project Budget Form Front

SECTION 1: DETAILED BUDGET
YEAR 2

Name of Applicant: Michael K. Buckland

SALARIES AND WAGES (PERMANENT STAFF)

NAME/TITLE	No.	Method of Computation	IMLS	Match	Total
TOTAL SALARIES AND WAGES			**$16,107**	**$23,601**	**$39,708**

SALARIES AND WAGES (TEMPORARY STAFF HIRED FOR PROEJCT)

NAME/TITLE	No.	Method of Computation	IMLS	Match	Total
Graduate Student Researchers	(2)	see Budget Attachment A	35,863		35,863
TOTAL SALARIES AND WAGES			**$35,863**	**$0**	**$35,863**

FRINGE BENEFITS

Rate	Salary Base	IMLS	Match	Total
TOTAL FRINGE BENEFITS		**$30,212**	**$4,012**	**$34,224**

CONSULTANT FEES

Name/Type of Consultant	Rate of Compensation (Daily or Hourly)	No. of Days or Hrs) on Project	IMLS	Match	Total
TOTAL CONSULTANT FEES					

TRAVEL

From/To	Number of Persons	Number of Days	Subsistence Costs	Transportation Costs	IMLS	Match	Total
West/East Coast Conf.	(4)	(20)	1,770	2,000	3,770		3,770
Research Meeting	(2)	(2)	50	180	230		230
TOTAL TRAVEL COSTS					**$4,000**	**$0**	**$4,000**

Application Form

1999 IMLS National Leadership Grants

Project Budget Form Back

SECTION 1 CONTINUED

YEAR 2

MATERIALS, SUPPLIES, AND EQUIPMENT

Item	Basis/Method of Cost Computation	IMLS	Match	Total
Copying, printing, computer supplies.		1,000	0	1,000
	TOTAL COST OF MATERIAL, SUPPLIES, & EQUIPMENT	**$1,000**	**$0**	**$1,000**

SERVICES

Item	Basis/Method of Cost Computation	IMLS	Match	Total
	TOTAL SERVICES			

OTHER

Item	Basis/Method of Cost Computation	IMLS	Match	Total
Effort contributed by Ex Libris		0	10,000	10,000
	TOTAL COST OF OTHER	**$0**	**$10,000**	**$10,000**
	TOTAL DIRECT PROJECT COSTS	**$87,182**	**$37,613**	**$124,795**

INDIRECT COSTS

Select either item A or B and complete C.

Applicant is using

A. an indirect cost rate which doesnot esceed 20% of direct costs

or

B. an indirect cost rate negotiated with a Federal agency (copy attached)

Department of Health & Human Services — Name of Federal Agency

3/25/97 — Effective Date of Agreement

C. Rate base(s) Amount(s)

50.4% of $ 59,596.00

Amount **$30,036**

TOTAL INDIRECT COSTS **$30,036**

1999 IMLS National Leadership Grants *Application Form* **6.9**

Project Budget Form

SECTION 2: SUMMARY BUDGET

Name of Applicant: Michael K. Buckland

IMPORTANT! READ INSTRUCTIONS ON PAGE 5.2 BEFORE PROCEEDING.

DIRECT COSTS

TOTAL		IMLS	MATCH	
	SALARIES AND WAGES	102,919	46,451	
	FRINGE BENEFITS	60,372	7,897	
	CONSULTANT FEES	—	—	
	TRAVEL	14,649	—	
	SUPPLIES & MATERIALS	2,000	10,771	
	SERVICES	—	—	
	OTHER	—	20,000	
TOTAL DIRECT COSTS		$ 179,940	$ 85,119	$
INDIRECT COSTS		$ 62,884	$ 37,471	$

TOTAL PROJECT COSTS $ 365,415

AMOUNT OF CASH—MATCH $ 122,590

AMOUNT OF IN-KIND CONTRIBUTIONS $ —

TOTAL AMOUNT OF COST SHARING (CASH AND IN-KIND CONTRIBUTIONS) $ 122,590

AMOUNT REQUESTED FROM IMLS $ 242,825

PERCENTAGE OF TOTAL PROJECT COSTS REQUESTED FROM IMLS (MAY NOT EXCEED 50% IF REQUEST IS ABOVE $250,000) 66.5 %

Have you received or requested funds for any of these project activities from another Federal agency? *(please circle one)* ▪ Yes ▪ (No)

If yes, name of agency _____ Date _____

Amount requested $ _____

1999 IMLS National Leadership Grants – Berkeley

BUDGET NARRATIVE

The project is for two years. The budget is the same for both years except as noted below.

Salaries and Wages of Permanent Staff.

Salaries and Wages of Temporary Staff. The Graduate Student Researchers are regular graduate students (ordinarily Ph.D. students) employed on an ad hoc basis for research projects. Because they are students, their employment is subject to strict University guidelines. They may not be employed more than 50% during the academic year. We have assessed the workload as two "full-time" Graduate Student Researchers, i.e. 50% during the academic year and 100% during the summers.

A 2% increase in salary rates is projected.

Fringe benefits have to follow University regulations: 17% for permanent employees. Graduate Student Researchers, being considered primarily students, receive minimal conventional fringe benefits (1.3% during the academic year; 3% during the summer), but they may be eligible for student fee remission and, if from out-of-state, Non-resident tuition fee remission.

Consultants. None.

Travel. Effective dissemination requires attendance and presentation at conferences and multiple conferences if both librarians, systems developers and digital library specialists are all to be reached effectively. We plan to make presentations concerning this project at the American Library Association Annual Meeting, the American Society for Information Science, the ACM Digital Libraries Conference and one other major library meeting, such as the Association for College and Research Libraries, the Library and Information Technology Association, the Public Libraries Association, the Special Libraries Association, or similar. Most conferences suitable for dissemination of this work are held on the East Coast or Midwest, which increases airfare and travel time and adds one hotel night for people living on the West Coast. The travel budget is based two long trips per annum (Midwest or East coast conferences) and two short trips (West coast) each year. Cost assumptions, based on recent experience, for each of the two long trips are: airfare $1,000,conference registration $260; hotel $425; and $200 per them = $1,885. For the two short trips:airfare $90 and $25 per diem = $115. This travel will be supplemented from other sources if available.

Equipment. A powerful Dell workstation ($10,771) is contributed as Match in the first year. All other equipment will be contributed as additional unbudgeted Match.

Supplies. Photocopying, printing, computer supplies (disks, toner, etc.), postage, miscellaneous office supplies: $ 1,000 per year.

Other. Ex Libris Ltd, based in Chicago, Tele Aviv, Hamburg, Luxemburg, and Hayes (U.K.) produces the ALEPH integrated library system used by more than 450 library systems in 38countries world-wide. ALEPH provides multi-lingual, multi-script and bi-directional language capabilities. Ex Libris is interested enough in the work in this proposal that they have assigned twenty-five working days per year of senior research and development staff to work us on the work proposed here. The expectation is that this will increase the speed at which the techniques that we developing will be incorporated into production software for library systems. We treat this as an in-kind matching contribution. We conservatively assess the value of this contribution as being $10,000(salary and fringe benefits)in U.S. terms, plus indirect costs. For more information see the Ex Libris (USA), Inc., web-site: http://www.exlibris-usa.com.

In-Direct Costs. The Indirect Cost Rate is the rate negotiated between the University of California, Berkeley, with the Federal government for Federal grants and contracts.